

AUDIO-ONLY BACKOFF IN AUDIO-VISUAL SPEECH RECOGNITION SYSTEM

Abstract of the Disclosure

Techniques for performing audio-visual speech recognition, with improved
5 recognition performance, in a degraded visual environment. For example, in one aspect
of the invention, a technique for use in accordance with an audio-visual speech
recognition system for improving a recognition performance thereof includes the
steps/operations of: (i) selecting between an acoustic-only data model and an
10 acoustic-visual data model based on a condition associated with a visual environment;
and (ii) decoding at least a portion of an input spoken utterance using the selected data
model. Advantageously, during periods of degraded visual conditions, the audio-visual
speech recognition system is able to decode (recognize) input speech data using
audio-only data, thus avoiding recognition inaccuracies that may result from performing
speech recognition based on acoustic-visual data models and degraded visual data.